160-TP-011-001

Data Migration Service (DMS) Development Plan for the ECS Project

Technical Paper--Not intended for formal review or government approval.

April 1996

Prepared Under Contract NAS5-60000

RESPONSIBLE ENGINEER

Tony Cetuk /s/ 4/24/96

Tony Cetuk Date
EOSDIS Core System Project

SUBMITTED BY

Joy Colucci /s/ 4/24/96

Joy Colucci, Science Office Manager Date
EOSDIS Core System Project

Hughes Information Technology Systems
Upper Marlboro, Maryland

This page intentionally left blank.

Abstract

This document describes the plan for development of the Configuration Items (CIs) and interfaces of the Data Migration Service (DMS) of the ECS. The plan identifies technical development factors required to implement the DMS, including issues related to the selection of the DMS architecture, internal/external interfaces and dependencies on COTS/ECS software packages. In addition, this document establishes the DMS functional requirements to clarify the presentation of the overall DMS system. The Development Plan is consistent with schedules and other technical efforts (e.g. Global DAAC Inventory) related to the data migration effort.

Keywords: DMS, Migration, Schedule, Configuration Item, Component, Toolkit, Integration and Test, HDF, architecture, requirement

This page intentionally left blank.

Contents

1. Introduction

2. Related Documentation

3. DMS Design Analysis

4. DMS Configuration Items

5. DMS Interfaces

Tables

Figures

Appendix A. DMS Functional Requirements

Abbreviations and Acronyms

1. Introduction

1.1 Purpose

This document describes the plan for development of the Configuration Items (CIs) and interfaces of the Data Migration Service (DMS) of the ECS. The plan identifies technical development factors required to implement the DMS, including issues related to the selection of the DMS architecture, internal/external interfaces and dependencies on COTS/ECS software packages. In addition, this document establishes the DMS functional requirements to clarify the presentation of the overall DMS system. The Development Plan is consistent with schedules and other technical efforts (e.g. Global DAAC Inventory) related to the data migration effort.

Note: This document does not provide an overview of the development environment, configuration management and the Project Instructions (PIs) that are applicable to the DMS streamlined development approach. These topics are covered in the DMS Software Development Plan, Document 160-TP-TBD-TBD.

1.2 Organization

The document is organized to describe the approach to develop the DMS:

Section 1.0 provides information concerning the purpose and organization of this document.

Section 2.0 provides a list of applicable documents which were used directly or indirectly in the preparation of this document.

Section 3.0 presents the driver analysis and design analysis for the DMS. Emphasis was placed on those areas that presented significant risk or cost impact. Findings from the analyses led to the selection of the DMS architecture.

Section 4.0 defines the CIs related to the Toolkit and Data Handling components of the DMS. The CIs include any dependencies on COTS/ECS software packages. The re-used CIs are mapped to their release A development schedule.

Section 5.0 establishes the system interfaces internal and external to the DMS, including the transfer mechanisms and sequencing of exchanges. All options of data exchange for each interface are described (e.g. polling with delivery record) as well as the conditions required for each option.

Appendix A lists the DMS Requirements defining the overall functional capabilities of the DMS. The requirements are allocated to specific components, according to the conceptual architecture, to aid and clarify the presentation of the overall DMS system.

Abbreviations and acronyms contains an alphabetized list of the definitions for abbreviations and acronyms used with this document

Questions regarding technical information contained within this paper should be addressed to the following ECS contacts:

- Tony Cetuk, HAIS, (301) 925-0421, tcetuk@eos.hitc.com
- Thomas Dopplick, HAIS, (301) 925-0333, tom@eos.hitc.com
- Herb Emerson, Loral, (301) 925-0619, hemerson@eos.hitc.com
- Larry Klein, ARC, (301) 925-0764, larry@eos.hitc.com
- Paul Roycraft, Loral, (301) 925-0623, proycraf@eos.hitc.com
- Ramachandran Suresh, HSTX, (301) 441-4092, suresh@hops.stx.com

1.3 Review and Approval

Questions concerning distribution or control of this document should be addressed to:

Data Management Office
The ECS Project Office
Hughes Information Technology Systems
1616 McCormick Drive
Upper Marlboro, MD 20785

1.4 Acknowledgments

The following people are acknowledged for having contributed to the writing of this document:

- Thomas Dopplick, HAIS
- Herb Emerson, Loral
- Larry Klein, ARC
- Jon Pals, HSTX
- Paul Roycraft, Loral
- Ramachandran Suresh, HSTX

2. Related Documentation

Figure 2-1 illustrates the relationships of data migration documents that pertain to design, development and test.

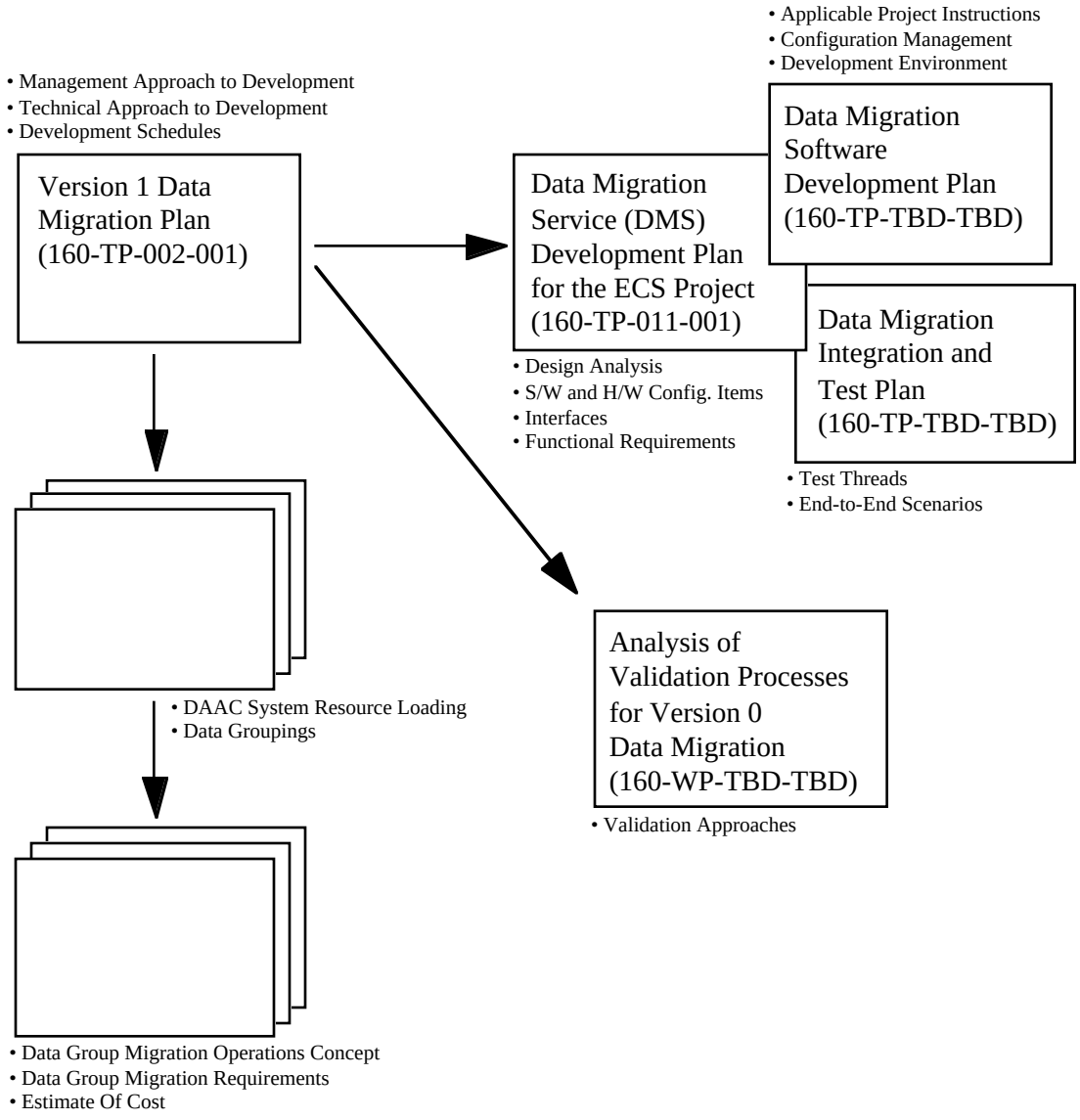

Figure 2-1. Data Migration Document Relationships

2.1 Applicable Documents

The following documents are referenced within this development plan or are directly applicable.

305-CD-004-001	Overview of Release A SDPS and CSMS System Specification for the ECS Project
305-CD-008-001	Release A SDPS Data Server Subsystem Design Specification for the ECS Project
305-CD-009-001	Release A Ingest Subsystem Design Specification for the ECS Project
308-CD-001-005	Software Development Plan for the ECS Project
160-TP-002-001	Version 1 Data Migration Plan
160-TP-006-001	Pilot Data Migration Report

2.2 Information Documents Not Referenced

The following documents, although not referenced herein and/or not directly applicable, do amplify and clarify the information presented in this document.

311-CD-002-004	Science Data Processing Segment (SDPS) Database Design and Database Specifications for the ECS Project
605-CD-001-003	ECS Operations Scenarios for the ECS Project: ECS Release A
194-815-SI4-001	SDP Toolkit Primer for the ECS Project [not available in hardcopy; see the ECS Data Handling System (EDHS) at http://edhs1.gsfc.nasa.gov/]
170-WP-002-001	Thoughts on HDF-EOS Metadata (White Paper for the ECS Project)
175-WP-001-001	HDF-EOS Primer for Version 1 EOSDIS (White Paper for the ECS Project

2.3 Draft Papers

Other papers relating to data formats are in process. Drafts of these papers were used in the preparation of this paper.

The HDF-EOS Swath Concept

The HDF-EOS Point Concept

The HDF-EOS Users Guide for the ECS Project

3. DMS Design Analysis

3.1 DMS Driver Analysis

The purpose of the DMS driver analysis was to identify the requirements and system needs that drive the design of the DMS. Emphasis was placed on those areas that presented significant risk or cost impact. The drivers were derived from several sources, including:

a. **Pilot Data Migration Study**

The Pilot Data Migration Study represents the findings from a coordinated study involving metadata comparisons between Version 0 and ECS for 34 data products and test granules obtained from the ASF, EDC, GSFC, JPL and LaRC DAACs. The investigation provided an opportunity to identify problems associated with data migration, including issues related to development of a generalized translator, collection and mapping of metadata and translation of data products into HDF-EOS. Details on the lessons learned are presented in the Pilot Data Migration Report, Document #160-TP-006-001.

b. **Conversations With DAAC Managers**

Technical exchange meetings are occurring with each DAAC Manager for the purpose of discussing the implications of data migration. In addition, the meetings provide an opportunity to elicit DAAC Manager feedback for identifying their specific concerns about V0 migration.

c. **DMS Requirements**

Appendix A lists the DMS Requirements defining the overall functional and performance capabilities of the DMS. The requirements are allocated to specific components, according to the conceptual architecture, to aid and clarify the presentation of the overall DMS system. This initial allocation of requirements is not to be considered restrictive of the final system architecture and design; this specification represents the baseline from which the DMS will evolve.

d. **Previous Migration Experience**

Migration efforts have been undertaken by private companies and government agencies for the purpose of translating legacy data and services into modern information systems. Many involve moving data from a mainframe environment to a relational database management system (RDBMS), involving creation of data repositories for the collection of metadata. The metadata repositories require mappings from legacy data attributes to their resultant migration data attributes. Occasionally, algorithms are required for deriving new attributes and transforming the data into a different format. Previous migration efforts provide additional insight into data migration issues and approaches.

The primary design drivers, along with their origin, are summarized in Table 3-1.

Table 3-1. DMS Key Drivers (1 of 2)

Design Driver	Origin	Description
Support an evolving baseline of migration products	DMS Requirements, Pilot Data Migration Study, Previous Migration Experience	The migration of specific data products will be spread out over Release A and Release B DMS operations. The DMS should provide an infrastructure that is unaffected by the incorporation of new products, allowing the system to operate while resource analysis continues. To accomplish this, the DMS should provide the following capability: a) Process control -- The migration of Version 0 data products will require different execution threads. In certain cases, one or more DMEs will be necessary for translating a product into an HDF-EOS data structure, creating ECS metadata or reformatting Version 0 data. Users will need the capability of creating and controlling these execution threads based on the product's migration requirements. b) Flexible metadata extraction -- Version 0 data products will contain metadata in different forms (e.g. the metadata may come in its own file or it may be within the data file). A flexible approach for extracting metadata is necessary to support the differences between the Version 0 data products.
Mapping of Version 0 attributes to ECS attributes	Pilot Data Migration Study, DMS Requirements	During resource analysis, many Version 0 metadata attributes can be mapped directly to ECS metadata attributes. Providing a mechanism for creating these associations and maintaining them in a repository will simplify the migration process.
Flexible architecture for supporting different QA/Validation schemes	Pilot Data Migration Study, Conversations with DAAC Managers (includes conversations with data experts)	Validation requirements will often vary from product to product. Validation can be performed in different ways, including checking the precision of numbers before and after translation, displaying and checking data visually and, in some cases, checking the contents of the data. The DMS architecture needs to support both automatic validation (software checking while the migration is taking place) and manual validation (off-line inspection of resultant data and metadata). In addition, providing the capability for the geographically distributed DAACs to be involved in the QA/Validation of their own data products will save time and effort in the migration process.

Table 3-1. DMS Key Drivers (2 of 2)

Design Driver	Origin	Description
Utilize standard and existing interfaces for data transfer	DMS Requirements, Conversations with DAAC Managers (includes conversations with operations staff)	In order to minimize the operational impact at the DAACs, the DMS should be designed to transfer data in a standard format and mechanism (i.e. as any other client that interacts with the SDPS Ingest service). In addition, the DMS should provide support for off-line (physical media) data transfer. Physical media will allow the DMS to operate in a standalone mode where migrated data can be stored for later insertion to the Data Server archive. In addition, physical media may provide benefits to network transfer with the Version 0 DAACs. For example, very large data products such as AVHRR have proven inherently difficult to transfer because of network bandwidth limitations.
Provide an early migration test bed to mitigate risks	Conversations with DAAC Managers, Previous Migration Experience	An early test bed will provide feedback on the DMS software infrastructure, with emphasis placed on high risk areas (e.g. interfaces) and areas with user interaction (e.g. graphical displays). In addition, a test bed would provide an early mechanism for performing sample granule migration and benchmark testing.

The DMS provides the capability for performing Version 0 data and metadata migration, but does not include the functionality for migration of services. Analysis is being performed to determine the existing services associated with Version 0 data. When appropriate, the same level of user services will be provided for the migrated data products. This may cause extensions to the SDPS Data Server to modify the baseline of available ECS services; however, it is beyond the scope of the DMS and will not be addressed in this document.

3.2 DMS Design Analysis

Based on the DMS drivers (see Section 3.1), several analyses were performed to address issues related to the DMS design and architecture. The findings of these investigations were based upon several sources, including:

- Conversations/Meetings with DAAC personnel
- Results from Pilot Data Migration Study
- Lessons learned from previous companies that performed data migration
- Review of existing COTS packages and public domain software
- Re-use of heritage code

Table 3-2 breaks down the primary design decisions related to the DMS. The overall details of the design and architecture are explained in Section 3.3.

Table 3-2. DMS Design Issues (1 of 2)

Design Driver	Decision	Rationale
Support an evolving baseline of migration products	Re-use SDPS Ingest software that includes templates for process control	For process control, several software approaches were investigated. The COTS package InConcert provided the user with graphical configuration of their process flows, but required user interaction for launching sequential jobs. Passport allowed a user to insert triggers for automatically launching DMEs, but the approach was rudimentary. By reusing software components from SDPS Ingest, a template strategy provides a GUI for specifying the manner to preprocess each data product. In addition, modifications can be made to each template without recompilation.
	Re-use SDPS Ingest software that includes configuration files for metadata extraction	A generalized translator approach was investigated for performing metadata extraction. Freeform (developed by NOAA's NGDC) worked well with simple data products (ASCII), but could not handle many common formats (e.g. CDF, netCDF, CEOS). The COTS package IMEX from Array Computing worked with simple formats like GIF, TIFF and Sun raster; however, many of the Version 0 data products for migration are in native formats. Likewise, IDL and SFDU Workbench were examined, but they do not handle an arbitrary native format. SDPS Ingest provided configuration files where a user specifies the input format and target format for specific data types. SDPS Ingest will handle different formats, including PVL, ODL and byte ordered. If a data product is not represented in these formats, the user has the option of either creating DMEs to reformat the data or modifying the SDPS Ingest software to include the necessary algorithms.
Mapping of Version 0 attributes to ECS attributes	Metadata Server with WWW access	Several options were considered in handling the mapping of Version 0 attributes to ECS attributes. One COTS product, Passport, offered adequate database form development, but did not provide the capability of creating HTML forms for WWW access. Two other COTS products, DBStar and Vality, offered the most functionality and compatibility, but were deemed unusable due to exceedingly high cost and considerable training. Freeform (developed by NOAA's NGDC) worked well with simple data products (ASCII), but could not handle many common formats (e.g. CDF, netCDF, CEOS). The decision to use an HTML based Sybase server was chosen because it provided WWW access, with customized code costing less than most of the COTS packages and able to handle different data formats.

Table 3-2. DMS Design Issues (2 of 2)

Design Driver	Decision	Rationale
Flexible architecture for supporting different QA/Validation schemes	Re-use SDPS Ingest s/w to support automatic validation	Before products are inserted into the data server archive, SDPS Ingest software performs automatic validation related to metadata existence, range checks and correctness of the product size. If a data product requires additional validation algorithms, the user has the option of either creating DMEs to perform the checks, or modifying the SDPS Ingest software to include the necessary algorithms. Re-using SPDS software for performing automatic validation provided both cost benefits and flexibility to add future enhancements.
	Provide off-line data examination tools for manual validation	For manual validation, data visualization tools will be provided to support off-line analysis. The validation approach will often vary from product to product, with the types of tools including EOSView, NCSA, public domain and COTS. In addition, the metadata server will include a WWW interface, allowing the external science community to inspect granule-level metadata mappings and collection-level metadata values before processing begins. Note that the final selection of the validation tools and strategy is still under review -- results and rationale will be published in the following white paper: Analysis of Validation Processes for Version 0 Data Migration (160-WP-TBD-TBD).
Utilize standard and existing interfaces for data transfer	Reuse SDPS Ingest software to include Polling with Delivery Record	For the transfer of data between the DAACs and the DMS, the polling with delivery record approach provides a standard ECS transfer protocol. Re-using software from SDPS provides the functionality related to the polling mechanism. With SDPS Ingest software already being selected for aiding process control and metadata extraction, there were cost benefits to utilizing it to also support data transfer.
	Reuse SDPS Data Server software to include physical media data transfer	SDPS Data Server provides the capability to perform data transfer to/from physical media. By reusing the software, the DMS could take advantage of the standard ECS transfer mechanisms, in addition to the cost benefit of incorporating existing software.
Provide an early migration test bed to mitigate risks	Build Metadata Server prototype; Build DMS test bed	Building an early version of the metadata server would provide feedback on the graphical interface necessary to associate the Version 0 granule-level metadata to the ECS granule-level metadata. Creating an early version of the DMS would support end-to-end testing of sample granules.

3.3 DMS Architecture

The design analysis from section 3.2 drove the DMS architecture that consists of three different elements: Data Migration Pre-processing System (DMPS), Data Migration Repository and Data Migration Executables (DMEs). The three architectural elements are presented in Figure 3.3-1, including their interactions between each other. All three components are described in detail in the following sections.

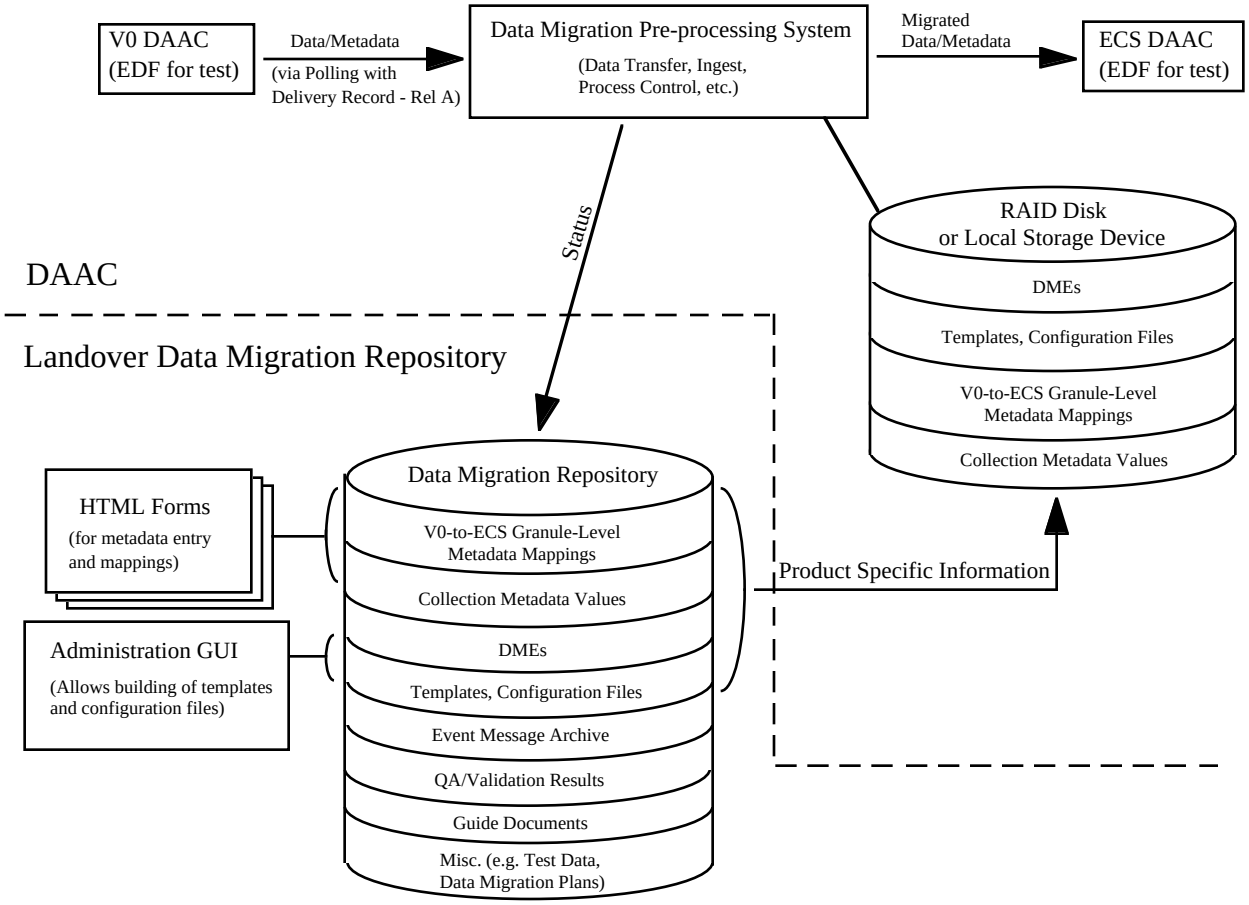

Figure 3.3-1. DMS High-Level Architecture

3.3.1 Data Migration Pre-processing System

The Data Migration Pre-processing System provides the infrastructure components required to support the migration process. Supported functions include transfer of data from Version 0, management of data flow and operations being performed on the data within the DMS, and transfer of the data to ECS for insertion into the SDPS Data Server Subsystem. The DMPS is comprised of a combination of hardware and software that incorporates significant reuse of Version 1 components. In addition to the reuse of experience and personnel in the design effort, the DMPS itself is built on a foundation of reused SDPS Ingest and Data Server components.

This approach helps reduce the complexity of the DMS development effort and mitigates the risk of developing and operating a separate migration string of hardware and software.

Transfer of data from Version 0 is accomplished electronically or via hard media through the use of two different reused SDPS computer software components (CSCs). Reuse of the Polling Ingest Client Interface CSC for electronic data transfer and the Peripherals CSC for data exchange via hard media provides reliable data transfer mechanisms that share basic functionality and operator interaction with those used in Version 1. Once data has entered the DMS through one of there mechanisms, reuse of the Ingest Request Processing Manager provides a template that instantiates other reused SDPS CSCs, as well as providing a means to control and moderate the flow of data through the DMS. The Ingest Data Preprocessing CSC provides the basic structure that supports the metadata extraction, data format conversion, and other functions being supplied by the DMS Toolkit components. A series of templates and configuration files are utilized to describe and control the processing functions that must be performed on each data type. A single InDataTypeTemplate will be configured for each Version 0 data type to be migrated by the DMS. An InFileTypeTemplate, describing the characteristics of the metadata and identifying the DMEs to be called to process the data, is prepared for each unique Version 0 file type. Templates are provided for byte ordered and Object Definition Language (ODL) files. Other templates may be required for the Version 0 file types, and may be added, as necessary. The third element is the InSourceMCF (Metadata Configuration File) that maps the source and target parameters.

Specific data handling component configuration items are discussed in Section 4.3. Additional details on the Data Handling component/Toolkit interface, including the schema for the data preprocessing templates and configuration file described in the previous paragraph, is provided in Section 5.2.

3.3.2 Landover Data Migration Repository

Before the migration process can begin, information must be collected pertaining to various characteristics of the Version 0 data. Analysis will be necessary to establish granule-level metadata mappings and collection metadata values. In addition, the requirements for migrating each data product may involve development of DMEs and process flows. A large assortment of information and processes will need to be obtained for each data product before their operational migration task can proceed.

The Data Migration Repository represents an intermediate data store for creating and maintaining the variety of information related to data migration. The Data Migration Repository is based on an RDBMS (Sybase) that is managed by programs and graphical interfaces. Many of the graphical interfaces will be WWW-based, allowing the DAACs, and the external science community as appropriate, to inspect product-specific information before migration begins. By providing an intermediate data store, off-line testing and validation can be performed on the metadata mappings, DMEs and migration procedures. Once fully tested, the product-specific information is ready to be moved to the DAACs to begin the operational migration process.

Table 3.3-1 summarizes the types of information that will be maintained in the Data Migration Repository.

Table 3.3-1. Information Types Contained In the Data Migration Repository

Information Type	Description
Data Migration Metadata Server (DMMS)	Version 0 to ECS granule-level metadata mappings and collection metadata values. Users will be provided with a WWW-based graphical interface for viewing and modifying the information.
Data Migration Information Server (DMIS)	Descriptive characteristics of the migrating data products, such as data format type, available read software and guide availability. This information is not collection or granule-level metadata, but is intended to assist the migration process by providing general information on each product. Users will be provided with a WWW-based graphical interface for viewing and modifying the information.
Templates and Configuration Files	Process flow descriptions for performing a product's data migration, including source/target metadata file formats.
DMEs	Executables necessary for performing specific migration functions, such as translation of native formatted data to HDF-EOS, metadata extraction and ODL generation. DMEs will include SDP Toolkit message handling and I/O as required.
QA/Validation Results	Results of migration data quality assessment.
Event Message Archive	History logs of informational messages generated during the migration process. Queries on message archives can provide summary information on the inventory of a DAAC's migrating product holdings, including information on each data product's migration history.
Guide Documents	Descriptive information on one or more data sets and related entities, containing details on the data content and potential usefulness for a specific application.
Test Data and Results	Sample test granules necessary for performing end-to-end testing at the EDF. Resultant products and event messages will also be retained for performing quality assessment.

3.3.3 DMEs

DMEs represent C or FORTRAN processes for performing specific migration functions, such as translation of native formatted data to HDF-EOS, metadata extraction and ODL generation. The DME architecture is based upon several subroutine libraries represented by the software usage hierarchy diagram in Figure 3.3-2. DMEs will also take advantage of constructs used in the ECS Planning and Data Processing System, such as Status Message Files and Metadata Configuration Files. By taking advantage of tested code to simplify the migration process, development and test cycles for building DMEs are shortened, thereby reducing overall life cycle costs.

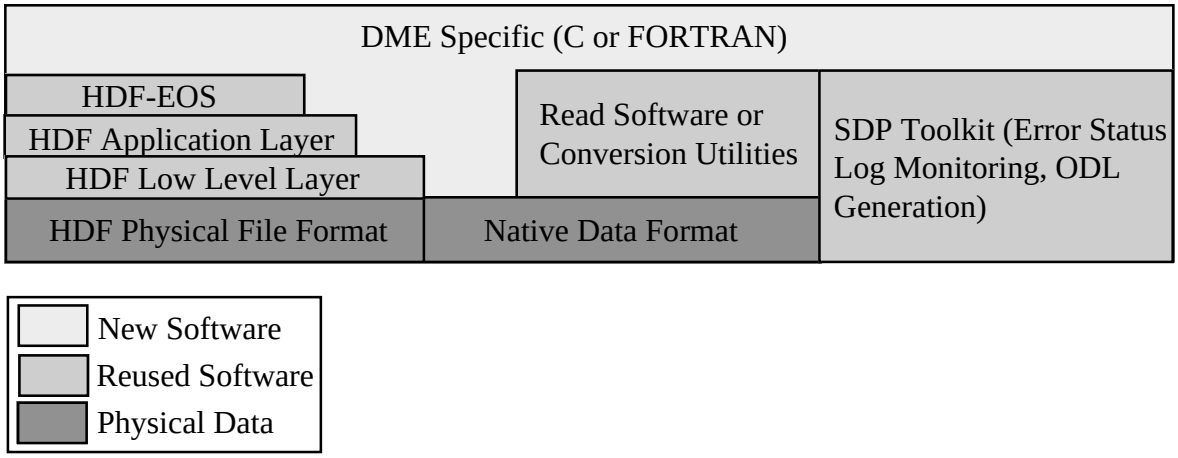

Figure 3.3-2. DME Software Architecture

All reused software libraries consist of callable C or FORTRAN routines. These libraries will provide DMEs with the following functionality:

- HDF Library - Creating and accessing NCSA HDF data structures, such as arrays, tables and annotations
- HDF-EOS Library - Extensions to the NCSA HDF library for creating and accessing three additional data structures: Point, Swath and Grid
- SDP Toolkit Library - Reporting of error/status messages and ODL generation for data ingest

This page intentionally left blank.

4. DMS Configuration Items

4.1 Overview

This section establishes the configuration items (CIs) involved in the implementation of the DMS. CIs represent an aggregation of software and/or hardware which can be individually tested and designated for configuration management. The DMS is logically divided into Toolkit and Data Handling CIs. The Toolkit components provide the means for performing metadata collection, data conversion to HDF-EOS and validation of migrated products. Data Handling components establish the primary DMS infrastructure, including data transfer, metadata extraction and process control.

4.2 Toolkit Configuration Items

The Toolkit CIs encompass all functionality related to development of the Data Migration Repository and DMEs. In addition, data visualization tools will be provided to support manual validation of migrated data at the DMS. Table 4.2-1 describes each of the Toolkit CIs.

Table 4.2-1. Toolkit Software Configuration Items (1 of 2)

Architectural Element	CSC	Description	Reuse Category
Data Migration Repository	DMMS Mapping Tool	Provides a WWW accessible graphical interface for setting up the mappings between the Version 0 and ECS granule-level metadata.	New
	DMMS Forms	Graphical database forms for displaying and updating collection metadata values (accessible via WWW).	New
	DMMS Procedures	Metadata Server Sybase procedures for performing database queries and updates based on user graphical selections.	New
	DMIS Forms	Graphical database forms for displaying and updating information pertaining to the DMIS (accessible via WWW).	New
	DMIS Procedures	DMIS Sybase procedures for performing database queries and updates based on user graphical selections.	New
	Data Extractor	Extracts granule-level metadata mappings, collection-level metadata values and other information from the Data Migration Repository for use in the operational DMS.	New

Table 4.2-1. Toolkit Software Configuration Items (2 of 2)

Architectural Element	CSC	Description	Reuse Category
DMEs	HDF Subroutine Library	Consists of callable C or FORTRAN routines for creating and accessing HDF data structures, such as arrays, tables and annotations.	Reuse
	HDF-EOS Extensions	Software extensions to the HDF subroutine library for creating and accessing three additional data structures: Point, Swath and Grid (C or FORTRAN).	Reuse
	SDP Toolkit Subroutine Library	Consists of callable C or FORTRAN routines for the reporting of error/status messages and to simplify PVL generation for data ingest.	Reuse
	DME Specific	Specialized software for performing specific migration functions, such as translation of native formatted data to HDF-EOS, metadata extraction and PVL generation	New
	Read Software	DAAC provided software utilities to simplify accessing native formatted data.	Reuse
	General Data Conversion Utilities	Basic conversion utilities that either partially or fully translate native data formats into HDF. For example, public domain software will be used for conversion from netCDF to HDF.	Reuse
DMPS	Data Validation Tools	Graphical visualization tools (i.e. EOSView, NCSA tools, public domain tools, COTS) for performing examination of migrated products. Note that the final selection of the validation tools and strategy is still under review -- results and rationale will be published in the following white paper: Analysis of Validation Processes for Version 0 Data Migration (160-WP-TBD-TBD).	Reuse

4.3 Data Handling Configuration Items

The DMS Data Handling CIs will be utilized to perform the data and metadata input and processing functions required to pre-process the Version 0 data to ECS standards before ingest into ECS. The DMS consists of multiple entities, with one string of equipment at the ECS Landover location and one at each of the DAACs where data is migrating during a specific ECS release. The DMS components are divided into h/w and s/w configuration items as described in the following sections.

4.3.1 Data Handling Hardware CI

DMS h/w will consist of components in the Landover EDF and at each of the DAACs. The Landover components will be used for s/w development, migration of sample granules, and I&T of the developed s/w and migrated granules. The components at the DAACs will be loaded with the developed Toolkit and Data Handling s/w, modified as necessary to conform to any DAAC-unique parameters, and otherwise prepared for the operational migration phase. This discussion focuses on the Release A configuration both in the EDF and at the Release A DAACs. In the EDF, the DMS will utilize hardware assigned to the Ingest subsystem development and I&T organizations. This h/w will be available on an as-scheduled basis for use by V0 DMS development and I&T personnel.

DMS h/w at the Release A DAAC sites also utilizes Ingest subsystem components. At LaRC, the DMS will consist of the backup Ingest SGI Challenge L and a TBD amount (estimate 30 GB initially) of RAID disk allocated to that host. This is also the configuration for the MSFC h/w that is being set up at the GSFC DAAC. The ingest configuration at GSFC is somewhat different for Release A, since the h/w in place at Release A is in support of early interface testing for the AM-1 mission. The GSFC DMS will utilize the SGI Indigo host for the V0 migration when it is not being used for interface testing. Peripheral devices (e.g., 8 mm tape drives) will also be available for use by the DMS, as a network connection providing access to other ECS subsystems and the Version 0 system at that DAAC.

There is currently an open issue regarding the MSFC DAAC ingest h/w located at GSFC. The backup ingest host in this string may also be available for use in migrating the GSFC Version 0 data, based on the degree of integration of the MSFC equipment into the GSFC DAAC configuration. The h/w and s/w configuration at each of the Release A DAACs for all ECS subsystems is documented in the Release A DAAC Design Specifications (GSFC:305-CD-014-001, LaRC:305-CD-015-001, MSFC:305-CD-016-001).

An additional open issue exists regarding the allocation of dedicated migration hardware at each of the DAACs. If approved, this dedicated hardware would become the primary hosts for performing migration specific tasks.

4.3.2 Data Handling Software Components

The DMS will utilize a combination of V0 migration Toolkit components and reused Ingest and Data Server subsystem components. Code development is divided into complete reuse, partial reuse, and new code. Table 4.3-1 describes the Ingest and Data Server Subsystem computer software components (CSCs) that are planned for use within the Data Handling component of the DMS.

4.4 COTS Software

Both the Data Handling and Toolkit components will make use of COTS software for the development environment, supporting the generation and modification of software. Additional COTS software is required to support integration and test activities. Table 4.4-1 summarizes the required COTS software.

Table 4.3-1. Ingest and Data Server Reuse Candidate CSCs

Architectural Element	CSC	Description	Reuse Category
DMPS	Ingest Session Manager	Provides a template that instantiates other client CSCs.	Reuse
	Polling Ingest Client Interface	Polls for data files or Delivery Record files in a specified location.	Reuse
	Ingest Request Processing	Moderates ingest processing steps	Reuse
	Ingest Data Preprocessing	Performs required preprocessing and interface with the Data Server for data insertion	Partial reuse
	Ingest Data Transfer	Transfers data from source to ECS staging space	Reuse
	Operator Ingest Interface	GUI screens allowing operations staff ingest of hard media, ongoing ingest request status monitoring, completed ingest request information viewing, ingest request controlling, and ingest threshold controlling	Partial reuse
	Service Clients	Provides client interface to ECS archives and the peripheral devices used for data ingest and distribution	Partial reuse
	Resource Management	Provides the capability to manage all resources of the Data Server Subsystem	Partial reuse
	Peripherals	Provides access to devices used for data ingest and distribution	Partial reuse
	Data Distribution	Prepares requested data objects for distribution on specified media or via network	Partial reuse
	Template Database	A configuration control tool to track, control, and log the migration progress.	New
	Template Editor	An enhanced GUI will be developed with the capability to create, store, and edit the library of templates required to support the large variety of V0 input, output, and conversion formats.	New
	Test Configuration Tools	GUI/database tools to track test data, to track test status for each data product, and to generate management reports on test and operational V0 data migration will be developed.	New
	Desktop	A GUI-based integrated desktop to interface between the new features and reuse code.	New

Table 4.4-1. COTS Software

COTS Support Area	COTS S/W Package	Description/function
Development	Builders Xcessory	GUI builder
	Clearcase	Configuration management tool
	C	Development language
	C++	Development language
	FORTRAN	Development language
	RogueWave	C++ class libraries
	RogueWave tools.h+	C++ class libraries
	Sybase	DBMS
	DCE	Support of Distributed Computing Environment
	OODCE	Object Oriented DCE
	SparcWorks	SUN development tools
	EPAK	GUI builder product
	Purify	Memory leak detector
	WABI	MS Office interface
	SparcWorks	Bundled SUN developer tools
	Netscape	WWW browser
Integration and Test	Xrunner	Test capture/playback tool
	Load Runner	Performance analysis tool

This page intentionally left blank.

5. DMS Interfaces

5.1 Overview

This section describes both the internal and external interfaces required to implement the DAAC DMPS, including the transfer mechanisms and sequencing of exchanges. The DMPS has one internal interface between the Toolkit and Data Handling components dealing with the retrieval/storage of data and metadata during the V0 migration process. For external interfaces, the DMPS is designed as any other client that interacts with the SDPS Ingest service. Running in the off-line mode, the DMPS will use the Data Handling component to retrieve/store information on physical media (e.g., 8mm).

5.2 Toolkit/Data Handling Interface

Part of the design of the DMPS is based on a relationship between the Toolkit components and the Data Handling components that extensively reuses elements of the ECS Ingest Subsystem currently under development. The interaction between the functions performed by the Toolkit and Data handling components is very similar to the interaction between the Ingest Data Preprocessing CSC and the Ingest Request Processing CSC. The DMPS will utilize elements of the Ingest request processing s/w to moderate and control the flow of data into and through the DMPS. The Toolkit components will augment portions of the ingest preprocessing s/w that, in the ECS design, performs functions including required data conversion, metadata extraction, and metadata existence and parameter range checks. The configuration of the Ingest Data Preprocessing CSC and the manner in which it interacts with the Ingest Request Processing CSC is documented in section 4.5.5 of the Release A Ingest Subsystem Design Specification (305-CD-009-001). Figure 5.2-1 depicts the relationship between the Data Handling Preprocessing elements and the Toolkit data processing and formatting tools. This is the interface that should be followed for the Toolkit to Data Handling component interface. Note that the update to this document for Release B CDR should be consulted to view possible refinements to this interface. The schema for the InDataTypeTemplate, InDataTypeTemplate, and InSourceMCF tables utilized by the preprocessing components is shown in Figure 5.2-2.

5.3 DAAC/DMPS Interfaces

The following sections describe the external interface between the DAAC and DMPS.

5.3.1 V0 DAAC/DMPS Interface

The DMPS will utilize standard Ingest subsystem ingest mechanisms for the transfer of L0 data. Both the LaRC and GSFC DAACs have selected the Polling with Delivery Record mechanism. Ingest s/w within the DMPS will initiate the transfer of data from the V0 DAAC system when it has been notified (via location of a Delivery Record at a specified polling location) that data is available for ingest. The handshaking/control messages and file transfer sequences related to the Polling with Delivery Record mechanism may be found in the ECS/EDOS ICD (Document number TBS).

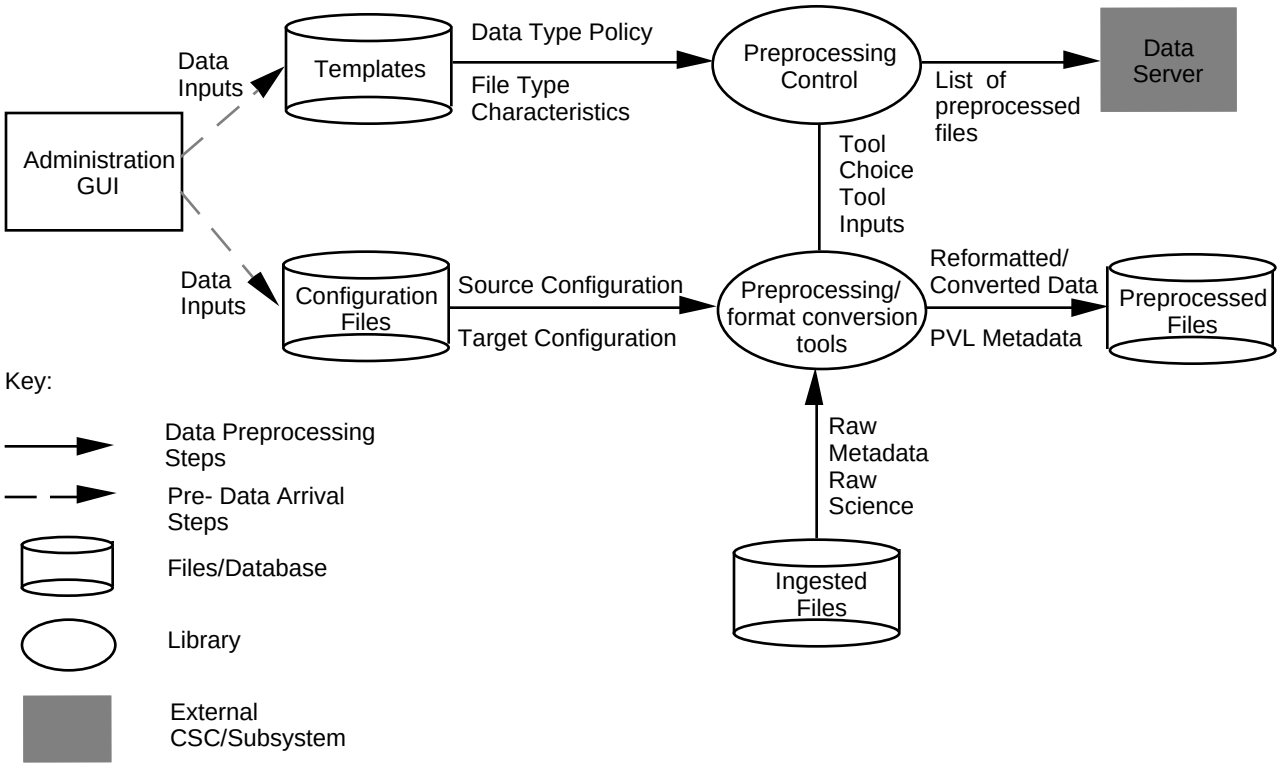

Figure 5.2-1. Data Handling Pre-processing CSC/Toolkit Interface in the DMPS

Off-line transfer of data from Version 0 to the DMPS will be accomplished via a hard media interface. The Ingest subsystem Hard Media interface mechanism provides a means for the DAAC to write V0 data to media (8 mm tape is preferred; 4 mm tape is acceptable) that is later ingested into the DMPS. Deliveries include at least 3 files consisting of delivery record, metadata, and one or more data files. The data files may be tar'd, at the discretion of the DAAC. Section 4.6 of the ICD Between ECS and Science Computing Facilities (SCF) (209-CD-005-004) defines the format and contents of the delivery record and metadata files.

5.3.2 DMPS/ECS Interface

The mechanism for transfer of data from the DMPS to ECS is similar to that used for the V0 DAAC to DMPS transfer. ECS Ingest will support both on-line (polling interface mechanism) and off-line (hard media ingest) transfer of data from the DMPS. Under nominal conditions, data that has been converted by the DMPS and is ready for ingest into ECS will be transferred utilizing the a polling with delivery record mechanism. The ingest s/w will be configured to poll a specified location for the presence of a delivery record file. When a delivery record is located, the ingest s/w will pull the data from the specified location. The hard media interface will be supported by Ingest s/w that reads the delivery record and ingests the corresponding metadata and data files. See the documents referenced in section 5.3.1 for a more detailed description of the data transfer message content and format for each of the supported ingest mechanisms.

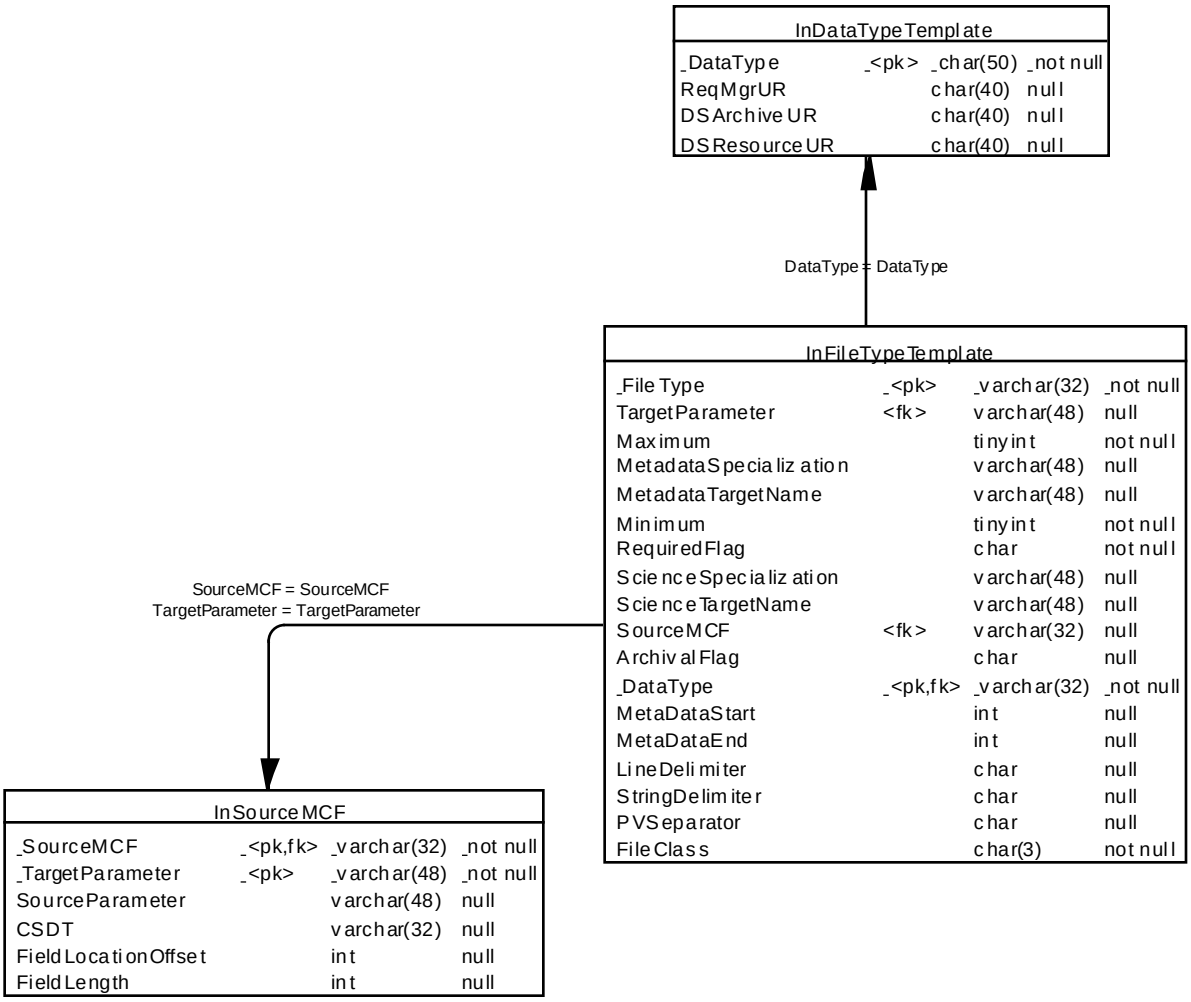

Figure 5.2-2. Data Preprocessing Table Schema

This page intentionally left blank.

Appendix A. DMS Functional Requirements

1.0 General

DMS-1000 The DMS shall have the capability to perform migration of V0 data products without interfering with the operational DAAC environment.

DMS-1010 The DMS shall be capable of operating in both an interactive mode and a batch mode.

DMS-1020 The DMS shall perform the migration of each Version 0 product based on the information and requirements specified in the product's corresponding Data Group Data Migration Plan.

DMS-1030 The DMS shall store and maintain information needed to perform data migration, including:

a. Migration software and procedures

b. Granule-level metadata mappings from Version 0 to ECS

c. User provided collection metadata values

d. Benchmark test procedures, test data and results

e. Data Migration Plans

f. DMS Operational Configuration

DMS-1040 The DMS shall notify system operators of conditions defined as requiring operator action.

DMS-1050 The DMS shall display informational and error messages to the system operators.

Note: Assumes the event messages will be sent to the MSS subsystem of CSMS.

2.0 Toolkit Components

DMS-2000 The DMS shall provide the capability to generate ECS granule-level metadata values based on:

a. Mappings from Version 0 metadata attributes

b. Algorithmically computing based on data product contents

DMS-2010 The DMS shall provide access to information related to the collection and validation of ECS metadata.

2.1 Metadata Server

DMS-2100 The DMS shall maintain and provide the capability to display an on-line metadata inventory containing information describing granule and collection level metadata.

DMS-2110 The DMS shall provide the capability to map Version 0 granule-level metadata attributes to the ECS granule- level metadata attributes (listed in SDPS Database Design and Database Schema Specifications for the ECS Project, CDRL #311-CD-002-004).

DMS-2120 The DMS shall maintain any product-specific metadata attributes that do not map to an ECS metadata attribute.

DMS-2130 The DMS shall provide the capability for a user to manually enter values for collection metadata.

DMS-2140 The DMS shall have the capability to interactively check collection metadata values for correct type and range.

DMS-2150 The DMS shall provide the capability to save metadata information established in prior user sessions for use in both current and future sessions.

DMS-2160 The DMS shall provide the capability to retrieve and display metadata information for use by DAAC personnel.

DMS-2170 The DMS shall restrict update of the metadata server to authorized users based on the user's access privileges.

DMS-2180 The DMS shall restrict QA/Validation of the collected metadata values to authorized users based on the user's access privileges.

DMS-2190 The DMS shall provide data base administration utilities for:

a. Modifying the database schema

b. Administration of user access control

c. On-line incremental backup

d. On-line recovery

e. Export/import of data

DMS-2200 The DMS shall maintain a log of all metadata update activity.

2.2 QA/Validation

DMS-2300 The DMS shall provide the distributed science community with the ability to validate:

a. Mappings from Version 0 granule-level metadata attributes to ECS granule-level metadata attributes

b. Manually entered values for collection level metadata

DMS-2310 The DMS shall have the capability to perform automatic QA/validation of a migrated product utilizing migration software as specified in the product's Data Migration Plan.

DMS-2320 The DMS shall have the mechanism for accepting product quality responses from the user community.

DMS-2330 The DMS shall provide benchmark test results for user validation.

DMS-2340 The DMS shall provide a data visualization tool to assist investigators in performing QA/validation of products migrated by the DMS.

DMS-2350 The DMS data visualization tool shall provide capabilities for image manipulation (e.g. pan, zoom, color, contrast).

DMS-2360 The DMS data visualization tool shall provide the capability to generate:

a. 2-D plots

b. 3-D plots

c. Contour plots

3.0 Data Handling Components

DMS-3000 The DMS shall be capable of transferring data with SDPS in the following manners:

a. Electronic communications network

b. Physical media (e.g. 8mm tape)

DMS-3010 The DMS shall be capable of complying with data transfer cancellation or delay notifications (TBR).

Note: This requirement is only necessary if the DAN approach is used for data transfer.

DMS-3020 The DMS shall notify the originating source of the need to retransmit data in the event of transmission difficulties.

DMS-3030 The DMS shall maintain data storage inventories defining the physical location of files that are part of the migration process.

3.1 Data Transfer Management

DMS-3100 The DMS shall be responsible for initiating the transfer of data and data transfer messages with SDPS.

DMS-3110 The DMS shall be capable of preparing Data Availability Notices for automatically transferring information with SDPS (in the manner described in section 4 of the ICD Between ECS and TSDIS, CDRL #209-CD-007-003).

DMS-3120 The DMS shall be capable of performing Polling with Delivery Record for automatically transferring product information with SDPS (in the manner described in the ICD Between ECS and EDOS, CDRL #TBD).

DMS-3130 The DMS shall provide authorized users the capability to manually control the data transfer process, including:

a. Initializing execution of data transfers

b. Suspending execution of data transfers

c. Resuming execution of a suspended task

d. Canceling execution of data transfers

Note: Suspend and resume are SDPS functions planned for release B.

DMS-3140 The DMS shall provide the user capability to monitor the ingest status of all product transfers with SDPS.

3.2 Physical Media

DMS-3200 The DMS shall provide its operations personnel with the capability to transfer information via physical media.

DMS-3210 The DMS shall have the capability to mount physical media via automated means.

DMS-3220 The DMS shall provide the capability to store data on 8mm tape.

3.3 Automated Data Checking

DMS-3300 The DMS shall be capable of validating all transfer messages, including checks for:

a. Presence of required fields

b. Metadata parameters out of range

c. Correctness of the product size

DMS-3310 The DMS shall generate status information indicating the success or failure of metadata entry and data consistency checks.

3.4 Migration Process Control

DMS-3400 The DMS shall control data migration based upon the information and migration requirements contained in a product's Data Migration Plan, including:

a. The list of migration tasks (i.e. processes) necessary to carry out a product's migration

b. The order in which the tasks need to be executed

c. The input data products and formats required

d. Processing time and other system resources required

DMS-3410 The DMS shall provide authorized users the capability to control the data migration process, including:

a. Initialize execution of tasks

b. Suspend execution of tasks

c. Resume execution of a suspended task

d. Cancel execution of tasks

Note: Suspend and resume are SDPS functions planned for release B.

DMS-3420 The DMS shall provide the user capability to monitor the processing status of all migrating products.

3.5 Audit Trail (Accounting)

DMS-3500 The DMS shall provide the capability to display a data product's migration history (TBR).

DMS-3510 The DMS shall maintain a processing log that accounts for all updates to data products and metadata during the migration process.

DMS-3520 The DMS shall have the capability to generate an inventory and status of its migrating product holdings.

DMS-3530 The DMS shall maintain a history log that accounts for all product transfers with SDPS.

4.0 DMS Testing

DMS-4000 The DMS shall provide a test and validation environment that is fully compatible with the operational migration environment.

DMS-4010 The DMS shall be capable of supporting test activities based on the information and requirements listed in the corresponding Data Migration Plan, including support of:

a. Test migration software and procedures

b. Sample Version 0 data/metadata

DMS-4020 The DMS shall have the capability to support analysis of migration test results.

DMS-4030 The DMS shall have the capability to validate required migration characteristics prior to the operational phase. These characteristics shall include:

a. Operational compatibility with DAAC Version 0 and Version 1 components

b. Required resources (e.g. algorithm software size)

c. Required data inputs/outputs

DMS-4040 The DMS shall have the capability to transfer validated migration software from the test environment to the operational environment.

Abbreviations and Acronyms

CI	Configuration Item
DAAC	Data Active Archive Center
DAN	Data Availability Notice
DME	Data Migration Executable
DMIS	Data Migration Information Server
DMMS	Data Migration Metadata Server
DMPS	Data Migration Pre-processing System
DMS	Data Migration Service
ECS	EOSDIS Core System
ODL	Object Definition Language
PI	Project Instructions
PVL	Parameter Value Language
SDPS	Science and Data Processing Segment
V0	Version 0

This page intentionally left blank.